DOCUMENT RESUME

ED 043 097 EA 003 033

TITLE	Technical Report of a Project To Develop Educational Cost-Effectiveness Models for New York State.
INSTITUTION	New York State Education Dept., Albany. Bureau of School Programs Evaluation.
PUB DATE	Mar 70
NOTE	34p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$1.80
DESCRIPTORS	*Academic Achievement, *Cost Effectiveness, *Data Processing, *Expenditures, *Factor Analysis, Information Systems, Statistical Analysis

ABSTRACT

This document reports on a study that explored the feasibility of using New York State Education Department data to develop multiple regression equations required for simulating the effects of alternative expenditure configurations on pupil achievement. Specific study objectives were: (1) to test the effectiveness of factor analysis as a procedure for reducing the number of variables required as independent terms in regression equations; (2) to develop a series of equations for predicting academic achievement, given pertinent information about conditions that might affect output; (3) to evaluate the desirability of using curvilinear, rather than linear, prediction equations; and (4) to explore the feasibility of using existing data to develop equations for predicting the effects of program changes on program outputs. The study produced eight equations for estimating various dimensions of school district output. Appendixes include a catalog of information elements, conclusions based on correlation data, and application of the models. (Author/LLR)

EDO 43097

TECHNICAL REPORT OF A PROJECT TO DEVELOP EDUCATIONAL COST-EFFECTIVENESS MODELS FOR NEW YORK STATE

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

The University of the State of New York
THE STATE EDUCATION DEPARTMENT
Bureau of School Programs Evaluation
Albany, New York 12224
March, 1970

EA 003 033

THE UNIVERSITY OF THE STATE OF NEW YORK

Regents of the University (with years when terms expire)

1984 Joseph W. McGovern, A.B., LL.B., L.H.D., LL.D., D.C.L.,
Chancellor - New York

1985 Everett J. Penny, B.C.S., D.C.S.,
Vice Chancellor - - - - - - - - - - - - - - - - - - - White Plains

1978 Alexander J. Allan, Jr., LL.D., Litt.D. - - - - - - - - - - Troy

1973 Charles W. Millard, Jr., A.B., LL.D., L.H.D. - - - - - - - Buffalo

1972 Carl H. Pforzheimer, Jr., A.B., M.B.A., D.C.S., H.H.D. - Purchase

1975 Edward M. M. Warburg, B.S., L.H.D. - - - - - - - - - - - New York

1977 Joseph T. King, LL.B. - - - - - - - - - - - - - - - - - - Queens

1974 Joseph C. Indelicato, M.D. - - - - - - - - - - - - - - - - Brooklyn

1976 Mrs. Helen B. Power, A.B., Litt.D., L.H.D. - - - - - - - - Rochester

1979 Francis W. McGinley, B.S., LL.B., LL.D. - - - - - - - - - - Glens Falls

1980 Max J. Rubin, LL.B., L.H.D. - - - - - - - - - - - - - - - - New York

1971 Kenneth B. Clark, A.B., M.S., Ph.D., Litt.D. - - - - - - - Hastings on Hudson

1982 Stephen K. Bailey, A.B., B.A., M.A., Ph.D., LL.D. - - - - - Syracuse

1983 Harold E. Newcomb, B.A. - - - - - - - - - - - - - - - - - - Owego

1981 Theodore M. Black, A.B. - - - - - - - - - - - - - - - - - - Sands Point

President of the University and Commissioner of Education
Ewald B. Nyquist

Executive Deputy Commissioner of Education
Gordon M. Ambach

Associate Commissioner for Research and Evaluation
Lorne H. Woollatt

Assistant Commissioner for Research and Evaluation
William D. Firman

Director, Division of Evaluation
Alan G. Robertson

Chief, Bureau of School Programs Evaluation
David J. Irvine

FOREWORD

The "Development of Education Cost-Effectiveness Models, Phase I" study was carried out by Abt Associates, Inc. under contract to the State Education Department. The study was intended to explore the feasibility of using data extracted from the Education Department's data files to develop multiple regression equations (models) required for simulating the effects on pupil achievement of alternative expenditure configurations. The study was one of a series focused on the development of regression analysis procedures for convert ing empirical data into information for the educational planner and decision maker. The problem of interpreting, validating, and increasing the practical utility of prediction equations will be dealt with in future studies.

Starting with 187 variables for about 500 New York State school districts, the present study: (a) identified internally correlated clusters of predictor variables, (b) identified internally correlated clusters of criterion variables, (c) revealed the degree of association between each cluster of predictor variables and a representative variable from each of the criterion clusters, (d) yielded correlations between most of the individual predictor variables and each of the representative criterion variables, and (e) produced a set of eight linear and eight curvilinear equations for estimating various dimensions of school district output.

This study would not have been possible without the generous support of the following individuals and units of the Education Depart-

ment: James R. Barnes, Associate, Bureau of Statistical Services; Ruth E. Callaghan, Director, Division of Electronic Data Processing; George Elmendorf, Supervisor, Division of Electronic Data Processing; Joseph A. Forte, Chief, Bureau of Educational Data Processing; David J. Irvine, Chief, Bureau of School Programs Evaluation; Alan G. Robertson, Director, Division of Evaluation; Alan D. Stewart, Associate, Bureau of School Programs Evaluation; John J. Stiglmeier, Director, Information Center on Education.

Finally, acknowledgment is due Neil Webre and Richard Anderson, both of Abt Associates, for their analysis of the data, and to David Tallcott, Consultant with the Bureau of School Programs Evaluation, for developing the extract data file and the associated documentation.

CONTENTS

		Page
I.	Background of the Problem	1
II.	Procedures	2
III.	Grouping and Indexing of Variables	3
	Predictor Variables Criterion Variables	
IV.	Description of Criterion Variables	7
V.	Development of Prediction Equations	9
	Predictive Effectiveness of Linear Models Predictive Effectiveness of Curvilinear Models	
VI.	Interpretation of the Equations	11
VII.	Application of the Equations	12
VIII.	Limitations	14
IX.	Conclusions	15
X.	Recommendations	15
APPENDIX A.	Catalog of Information Elements	17
APPENDIX B.	Conclusions Based on Correlation Data	21
APPENDIX C.	Application of the Models	25

Background of the Problem

The New York State Education Department regularly collects a variety of data on many facets of elementary and secondary education in the State. The data constitute a potentially valuable source of management information. To more fully utilize this data, it is necessary to develop computer programs for economically and rapidly selecting and processing a wide range of data.

One of the objectives of the study was to test the effectiveness of factor analysis as a procedure for reducing the number of variables required as independent (predictor) terms in regression equations. A second objective was to develop a series of equations for predicting academic achievement (school district outputs), given pertinent information about conditions which might affect output. A third objective was to evaluate the desirability of using curvilinear, rather than linear, prediction equations. Finally, the study was intended to explore the feasibility of using existing data to develop equations for predicting the effects of program changes on program outputs.

Procedures

Virtually all of the machine-readable data available within the State Education Department were prepared for possible use in the project. The data are summarized in the "Catalog of Information Elements" (Appendix A). A number of specific steps were involved in the analysis

1. From the initial pool of variables (Appendix A), 187 variables were selected for analysis. Of these, 87 were potential predictor variables and 100 were potential criterion variables.
2. Variables, for which data for as many as one-third of the school districts in the sample were unavailable, were eliminated, leaving 137 variables. Of these, 68 were classified as predictor variables and 69 were classified as criterion variables.
3. Intercorrelations of the 137 variables were computed.
4. The 68 predictor variables were factor analyzed, resulting in 20 predictor factors (groups of related predictor variables). An index value was computed to represent each predictor factor in subsequent analyses.
5. The 69 criterion variables were factor analyzed, resulting in 13 criterion clusters. A "typical" variable within each cluster was selected to represent that cluster in subsequent analyses.
6. A multiple linear regression analysis was performed using the 20 predictor indices as independent variables, and the 13 typical criteria as dependent variables, one at a time. Five predictor variables were subsequently eliminated because they provided no useful prediction of any of the criteria. Five typical criterion variables were subsequently eliminated

because they were not predicted at a useable level (F=1.85) by the predictor indices. Fifteen predictor indices and eight typical criterion variables remained.

7. The analyses performed in Step #6 resulted in eight equations, one for each of eight typical criterion variables. Regression correlation coefficients (R), and coefficients of determination (R^2) were computed for each equation.
8. Twenty-seven interaction terms (consisting of combinations of the 15 predictor indices) were included along with the 15 index variables to form eight curvilinear equations. The inclusion of these interaction terms resulted in a substantial increase in the level of prediction for the eight new equations.
9. The final result was a set of eight curvilinear multiple regression equations. These equations were regarded as prototypes of equations which could be used to calculate expected achievement levels in eight subject areas for any school system comparable to the systems from which the data were obtained.

Grouping and Indexing of Variables

The methods by which the number of variables is reduced should maximize the ratio of relevant to irrelevant data used in the final equations. In this study, factor analysis was used to define relatively independent clusters of predictor and criterion clusters. The result of this procedure is described in this section.

Predictor Variables. Sixty-eight predictors were selected from a pool of 187 raw variables to improve the efficiency of

prediction equations by further reducing the number of predictor variables. The reduction was accomplished by identifying (through factor analysis) clusters of predictor variables which were most highly intercorrelated. Each cluster was represented by an index variable. The indexes and their component raw variables are listed below:

Index	Weight	Component Predictor Variables
ENROLLMENT STABILITY	+.593	x proportion of schools with guidance services
	+.583	x language labs per pupil
	+.441	x income per pupil from BOCES
	+.429	x revenues from State sources
	+.277	x proportion change in ninth-grade enrollment, 4 years
	+.254	x expenditures per pupil for texts, K-6
PUPIL POPULATION	+.982	x total enrollment
	+.979	x secondary enrollment
	+.973	x elementary enrollment
	+.951	x number of schools
	-.604	x community type index
	+.410	x expenditure per elementary pupil for prekindergarten
	-.353	x expenditure per pupil under BOCES
SECONDARY INSTRUCTION	+.869	x expenditures per pupil for salaries, 7-12
	+.696	x expenditures per pupil for texts, 7-12
	+.578	x expenditures per pupil for interscholastic athletics
	+.495	x business training rooms per pupil
	+.465	x science classrooms per pupil affected

Index	Weight	Component Predictor Variables
MAJORITY ETHNICITY	-1.000 x	proportion of Negro secondary pupils
	+.983 x	proportion of white secondary pupils
	-.860 x	proportion of Negro elementary pupils½
	+.759 x	proportion of white elementary pupils
COMMUNITY PROSPERITY	+.929 x	full valuation per pupil
	+.926 x	total property tax per pupil
	+.893 x	total instructional expenditures, per pupil, day school
	+.884 x	expenditure subtotal for teaching, per pupil
	+.505 x	expenditure subtotal for cocurricular activities, per pupil
	+.503 x	salaries per pupil, K-6
	-.471 x	revenue per pupil from State sources
	+.413 x	expenditure subtotal for guidance, per pupil
	+.363 x	expenditures per pupil for central administration
	+.243 x	revenue from other districts per pupil
	+.209 x	expenditures for texts per pupil, K-6
ELEMENTARY FACULTY	+.808 x	tuition income per pupil
	+.691 x	expenditure per pupil for guidance
	+.619 x	salaries for substitutes, per pupil
	+.612 x	salaries per pupil, K-6
	+.475 x	total expenditures per pupil for central administration
	+.390 x	revenue from other districts per pupil
	+.224 x	equipment expenditure per pupil
SERVICES	+.718 x	proportion of schools with psychological services
	+.703 x	proportion of schools with health services
	+.407 x	proportion of schools with social work services
	+.381 x	proportion of schools with miscellaneous innovations
	+.371 x	proportion of schools with regional programs
MINORITY ETHNICITY	+.839 x	proportion of Puerto Rican elementary pupils
	+.823 x	proportion of Puerto Rican secondary pupils
SECONDARY NONGRADED	+.817 x	proportion of high schools with non-graded program
	+.805 x	proportion of junior highs with non-graded program

Index	Weight	Component Predictor Variables
INNOVATIVE-NESS	+.615 x	proportion of schools with State or national innovations
	+.600 x	proportion of schools with consultants
	+.585 x	proportion of schools with local innovations
	+.557 x	proportion of schools with miscellaneous innovations
	+.500 x	proportion of schools with independent study
	+.405 x	proportion of schools with instructional television
	+.397 x	proportion of elementary schools with nongraded program
	+.342 x	proportion of schools with regional programs
SCHOOL SIZE	+.610 x	pupils per school
	+.282 x	proportion change in 9th grade enrollment, 4 years
FEDERAL INTERVENTION	+.786 x	federal revenues per pupil
	+.696 x	revenue per pupil from NDEA III
	+.395 x	revenue from other districts per pupil
	+.297 x	expenditures per pupil for equipment
ACADEMIC FACILITIES	+.793 x	shop facilities per pupil
	+.700 x	libraries per pupil
	+.622 x	professionals per pupil
	+.589 x	science classrooms per pupil
	+.306 x	language labs per pupil
NONACADEMIC FACILITIES	+.892 x	physical education facilities per student affected
	+.836 x	cafeterias per pupil
	+.650 x	total expenditures per pupil for instruction, special school
	+.441 x	business training rooms per pupil affected
UNCROWDED-NESS	+1.00 x	regular classrooms/pupil

Description of Criterion Variables

From the pool of 187 raw variables, 69 criterion variables were selected. Factor analysis was employed to help identify internally correlated clusters of criterion variables. Eight clusters were identified. It was decided not to develop index variables to represent each of these clusters since this would have resulted in indices which were not readily interpretable. Instead, a single variable in each cluster was allowed to represent the cluster. Eight criterion variables (representing each of the eight clusters) and the variables within their respective clusters are listed below:

Weight		Criterion (Output) Variables
1.000	x	PROPORTION LEAVING TO 2-YEAR SCHOOLS
.336	x	Proportion leaving to 4-year schools
.266	x	Proportion leaving to employment
.230	x	Proportion leaving to military
.217	x	Proportion leaving to other schools
1.000	x	PROPORTION PASSING REGENTS EXAM IN BOOKKEEPING
.527	x	Proportion eligible to take Regents exam in bookkeeping
1.000	x	PROPORTION PASSING REGENTS EXAM IN BUSINESS LAW
.332	x	Proportion eligible to take Regents exam in business law
.229	x	Proportion passing Regents exam in typing
1.000	x	PROPORTION PASSING REGENTS EXAM IN SHORTHAND
.592	x	Proportion eligible to take Regents exam in shorthand
.159	x	Proportion leaving to employment

Weight		Criterion (Output) Variables (con't.)
1.000	x	PROPORTION PASSING REGENTS EXAM IN LATIN
.537	x	Proportion eligible to take Regents exam in Latin
1.000	x	PROPORTION PASSING REGENTS EXAM IN PHYSICS
.382	x	Proportion eligible to take Regents exam in physics
.366	x	Proportion passing Regents exam in 11th grade math
.169	x	Proportion passing Regents exam in 10th grade math
1.000	x	MEAN OF THIRD-GRADE PUPIL EVALUATION PROJECT (PEP) READING SCORES
.795	x	Mean of third grade PEP arithmetic scores
.635	x	Mean of sixth grade PEP reading scores
.447	x	Mean of sixth grade PEP arithmetic scores
.437	x	Mean of first grade PEP scores (.437)
-.435	x	Standard deviation of third grade PEP arithmetic scores
-.425	x	Standard deviation of third grade PEP reading scores
-.300	x	Standard deviation of sixth grade PEP reading scores
-.244	x	Standard deviation of first grade PEP scores
1.000	x	MEAN OF NINTH GRADE PEP ARITHMETIC SCORES
-.820	x	Standard deviation of ninth grade PEP arithmetic scores

Development of Prediction Equations

A principal objective of this study was to develop a set of transitional equations of a type which could be used to estimate the output values of real or hypothetical input changes made in particular school systems. Two sets of equations were developed. One set was based on the hypothesis that the criterion variables could be adequately predicted by simple linear combinations of predictor variables. The second set of equations involved the hypothesis that certain criterion variables are better predicted by including selected interaction terms (multiplication of one independent variable by another) in addition to the original variables. The two hypotheses were described as "linear" and "curvilinear" respectively.

Predictive Effectiveness of Linear Models. The eight equations for estimating school system outputs from a linear combination of predictor variables yielded the following correlation coefficients:

Dependent Variable	Multiple Correlation (R)	Multiple Determination (R^2)
1. Proportion leaving to 2-year school	.465	.216
2. Proportion passing Regents exam in biology:	.416	.173
3. Proportion passing Regents exam in business law:	.298	.089
4. Proportion passing Regents exam in shorthand:	.482	.233
5. Proportion passing Regents exam in Latin:	.428	.183
6. Proportion passing Regents exam in physics:	.350	.123
7. Mean of third-grade PEP reading scores:	.458	.210
8. Mean of ninth-grade PEP arithmetic scores:	.409	.167

Predictive Effectiveness of Curvilinear Models. Twenty-seven interaction terms were introduced as predictors in addition to the 15 original predictor terms. The recent literature on education for the disadvantaged provided evidence for hypothesizing interactions of background variables (e.g. "community prosperity," "population," "ethnicity," and "school size") with manipulable inputs such as "academic facilities," "innovativeness," and "Federal intervention." (Cf., for example, Moynihan, "Sources of Resistance to the Coleman Report," Harvard Educational Review, Vol. 38, No. 1 (Winter, 1968), p. 33.) The interaction terms were products of background indices and other (manipulable) input indexes. The eight resulting curvilinear prediction equations produced the following coefficients:

Dependent Variables	Multiple Correlation (R)	Multiple Determination (R^2)
1. Proportion leaving to 2-year schools:	.558	.311
2. Proportion passing Regents exam in biology	.480	.231
3. Proportion passing Regents exam in business law:	.389	.151
4. Proportion passing Regents exam in shorthand:	.560	.313
5. Proportion passing Regents exam in latin:	.483	.234
6. Proportion passing Regents exam in physics:	.448	.201
7. Mean of third-grade PEP reading scores:	.526	.276
8. Mean of ninth-grade PEP arithmetic scores:	.515	.266

The predictive efficiency (R^2) of the curvilinear models ranged from .151 to .313. The increases in R^2 ranged from 30 to 70 percent over the R^2s obtained using the linear models. Use of the interaction terms appeared to be most effective for increasing the level of prediction of the ninth-grade PEP arithmetic scores (equation 8) and least effective for increasing prediction of proportion passing Regents exam in business law (equation 3).

Interpretation of the Equations

Regression equations such as those developed in this study are useful tools if they are interpreted judiciously. However, it is essential to distinguish between what a regression equation does and does not do. An educator who wishes to use such equations will provide the equations with various input (predictor) values, asking the equations, in effect, what those configurations imply for the outputs. When the user has succeeded in adjusting the input levels so that the ratio of costs-to-effectiveness is maximized he needs to know how confident he can be that the input changes, if actually made, would produce the predicted results. Suppose, for example, that one proposed to provide two new business education rooms in a schoo' district, and that the appropriate equation "predicted" that the consequence would be a 5 percent increase in proportion of students passing the Regents exam in shorthand 1 year later. In interpreting this result the user should know that in the strictest sense the model is making the following assertion: "If the two proposed rooms are added, the district will thereby come to resemble more closely districts which have had, in the past, 5 percent higher success rate on that test than the district's present inputs suggest." In making this assertion, the model takes the following into account:

1) the present inputs to the district (before the proposed change);

2) the proposed change; and

3) the information its parameters contain about the past associations of inputs and outputs in the school district population.

Note that the equation compares its estimate of postchange output with its estimate of prechange output, not with the real-world value of prechange output. It thus "predicts" directions and relative magnitudes of change, not absolute values of outputs.

A regression equation has, of course, no direct information about the future. It "knows" only what it has been told about the past and the present. It therefore "predicts" in only a limited sense: it provides the values of output variables most likely to be associated with given present or past values of input variables, based on information summarized in its internal parameter matrices. The insights provided by an equation become assertions about the future as the human client makes the common presumption of uniformity of past, present, and future: if other things remain as they were in the past, we expect thus-and-so in the future. The user should exercise special care in interpreting "predictions" from a regression equation if he has any reason to suspect that important changes have taken place in the real-world system on which the equation is based. If past and future differ sharply, the past may be a misleading guide to the future.

Application of the Equations

If the equations developed in this study were programmed for a computer, the user would simply provide values for the 60

raw input variables, whereupon the computer would perform the calculations required by the prediction equations. For a particular school district, or school building, these calculations would yield estimated values for the eight types of outputs, together with estimates of the reliability of these values.

The user could then provide new data--data about anticipated or hypothetical changes. Calculations could then be performed on the data to generate new output estimates. The user could propose any number of changes and the computer, in conjunction with the appropriate equations, would calculate the estimated outputs associated with each change. The user then could decide which change or set of feasible changes would be likely to produce the most satisfactory results. An example of the functioning of an equation is presented in Appendix C. A prediction equation might also be used as a means of calculating a discrepancy score for a school program or school district. For example, the difference between the actual output score for a school district and its expected output score as computed using a prediction equation could provide information that should help to identify possible problems that might prevent a program from having its maximum effectiveness.

Limitations

Although it would have been feasible to apply the equations developed in this study to help supply management-relevant information, it was felt that their usefulness was limited and that other equations should be developed and tested before being put into service. Problems associated with the existing equations are listed below:

Missing data. Compatibility problems involving the contractor's tape-handling facilities resulted in the loss of the Basic Educational Data System (BEDS) Personal file and the School Finance file for 1965-66. Because of the theoretical importance of these data for prediction purposes it was felt that their omission would limit the value of any equations which were developed without these data.

Omission of "strong" predictor variables. Due to an oversight achievement data collected in the lower grades (initial status data) were not used in developing the regression equations. Furthermore, intelligence and socioeconomic data were not available. Since these are theoretically important predictors of subsequent levels of achievement their omission from the equations resulted in an unnecessarily high proportion unspecified variance ("error").

Classification of Variables. The method of clustering the input and output variables by factor analysis resulted in groupings that were relatively homogeneous in a statistical sense. But these groupings were not necessarily meaningful to an administrator. Rather than identifying criterion (output) variables on the basis of their statistical "groupiness" it would probably have been more useful to identify output variables of interest to the administrator and to develop equations using these as the criterion variables.

Conclusions

The present study provided a partial test of the feasibility of using data from the Department's data bank for the purpose of developing equations for generating performance information. While the equations themselves were not regarded as sufficiently error-free to warrant direct application, the level of prediction obtained suggested that the data could be used to develop more efficient and effective prediction equations.

The methodology for producing equations was conventional in conception and it resulted in equations that were not outstanding in terms of predictive effectiveness. It is felt that better methods for extracting, grouping, compressing, and configuring predictor variables can be developed. The present study provided valuable insights which will be useful in future development.

The study was focused primarily on methodological problems. However, the correlations observed between various input and output variables were useful byproducts of the study. These correlations revealed intervariable relationships that previously might not have been apparent. It would not be advisable, of course, to base program decisions on these findings.

Recommendations

The procedures for isolating output (criterion) variables should be revised. The use of factor analysis as a technique for identifying groups of output variables resulted in groupings that frequently were not useful from the perspective of school decision makers. Often it is desired to predict how conditions will affect a specific output. In such cases, only the output of interest should be quantified and used as the criterion variable.

In cases where it is desired to study effects on a cluster of output variables, it would be more useful to the decision maker if these outputs were grouped on a conceptually meaningful basis rather than on the basis of the statistical covariances of the various outputs.

The procedures used for grouping input (predictor) variables should be revised. The factor analytic approach resulted in grouping input variables on the basis of their statistical covariance, not on the basis of their logical relatedness vis-a-vis the output (criterion) variable. It is recommended that future attempts to group predictor variables utilize methods which are oriented toward: (a) maximizing the prediction of a specified output variable, and (b) maximizing the interpretability of the effects of the input variables.

The retrieval of data from the data bank should be simplified. Considerable effort was required to obtain data needed to develop the prediction equations. Programs should be developed which would facilitate rapid production of an extract file of the desired data for specified schools or districts.

APPENDIX A

CATALOG OF INFORMATION ELEMENTS

Data Included in the Basic Educational Data System*

In essence, BEDS information ties a group of students taking a particular course to a teacher, to a school, to a school district. Specifically, the following data are available:

Students

- Number of students by course
- Grade level of students by course
- Ability level of students by course
- Racial composition of students by grade within school
- Distribution of high school graduates

  to college, in and out of New York State
  to employment
  to military service
  to other

Professional Staff (All)

- Sex
- Marital status
- Degree status
- Occupation prior year
- Location of occupation prior year
- Type of appointment
- Educational experience

  for each assignment
  in present school district
  in public education
  total educational experience

- Number of months employed
- Percent of time employed
- Salary
- Age

---

*This list was prepared by the Information Center on Education. The list reflects the type of data which were available at the time of this study. Due to equipment problems encountered by the contractor, personnel data were not used in this study. Also, only district-level data were used in this study.

BASIC EDUCATIONAL DATA SYSTEM (continued)

Professional Staff (Classroom Teachers)

- Teaching schedule, by assignment
- Certification for each assignment
- Time spent in each assignment
- Number of students in each assignment
- Ability level of students in each assignment
- Grade level of students in each assignment

Professional Staff (Nonclassroom)

- Assignments
- Certification for each assignment
- Grade level for each assignment
- Number of public school buildings for each assignment
- Number of public school students for each assignment
- Number of nonpublic school buildings for each assignment
- Number of nonpublic school students for each assignment
- Time spent in each assignment

Curriculum

- Course offerings
- Registration by course
- Grade level
- Type of pupils
- Time spent in each course

School Characteristics

- Number and racial composition of

  principals
  classroom teachers
  other instructional staff

- Enrollment and ethnic composition by grade
- Number and length of class periods
- Length of school day
- Holding power
- Number of dropouts
- Age of building
- Age of additions
- Number of rooms and facilities--32 categories
- Presence of selected special programs and facilities

School District Characteristics

- Number of instruction rooms abandoned, completed, available
- Pupils on curtailed session

School District Characteristics (continued)

- Number of professional and nonprofessional staff

  teacher aides
  secretaries, typists, and clerks
  maintenance workers and custodians
  bus drivers and mechanics
  school lunch workers

FINANCIAL DATA

For each school district, detailed revenue and expenditure data.

ATTENDANCE DATA

By grade groups (K, 1-6, 7-12) within school district.

- Average daily attendance
- Weighted average daily attendance
- Possible average daily attendance

CENSUS DATA

Census of minors, by age and sex

Minors on census attending school

- In public school in home district
- In public school away from home district
- In nonpublic school in home district
- In nonpublic school away from home district

Census of migrant children

- Number who resided in district in prior 12 months
- Number who were enrolled in public schools

Census of handicapped children

- Sex
- Age
- Types of handicap
- Latest examination
- IQ score, if mentally retarded
- Type of school attendance

NONPUBLIC SCHOOL DATA

Students

- Enrollment by grade
- Residence
- Distribution of high school graduates

Professional Staff

- Number by level and function

Curriculum

- Types of curriculum offered

School Characteristics

- Number of instruction rooms
- Presence of selected special programs and facilities

PUPIL EVALUATION PROGRAM

For each public and nonpublic school, test results as follows:

- Grade 1 - Readiness
- Grade 3 - Reading and arithmetic achievement
- Grade 6 - Reading and arithmetic achievement
- Grade 9 - Reading and arithmetic competence

REGENTS EXAMINATION RESULTS

For each public and nonpublic school, by subject area

- Number of students taking exam
- Number of students passing exam

REGENTS SCHOLARSHIP EXAMINATION

Test scores are available, by school, for approximately 75 percent of the public and nonpublic high school seniors who annually take this examination of college aptitude and high school achievement.

APPENDIX B

CONCLUSIONS BASED ON CORRELATION DATA

Correlational data were obtained for all of the variables used in the present study. Although it is difficult to draw useful conclusions from correlation coefficients, tables 1 and 2 are presented for the benefit of those who may wish to use the data for generating specific hypotheses. For example, the following observations in table 1 appear noteworthy:

1. Secondary expenditures were <u>negatively</u> related to all categories of achievement.
2. Community prosperity was <u>negatively</u> related to all but two categories of achievement (3rd grade reading and 9th grade arithmetic).
3. The output categories which correlated most highly with inputs were as follows (from most to least correlated): 3rd grade reading, proportion leaving to 2-year colleges, business law, shorthand, physics, Latin, 9th grade arithmetic, bookkeeping.

TABLE 1.

CORRELATIONS OF SCHOOL DISTRICT INPUT INDEX VARIABLES
WITH SCHOOL DISTRICT OUTPUT VARIABLES
(N=636)

	Output Variables							
Input Indices	% Leaving to 2-Year Colleges	% Passing Bookkeeping Regents	% Passing Business Law Regents	% Passing Shorthand Regents	% Passing Latin Regents	% Passing Physics Regents	Mean 3rd Grade Reading	Mean 9th Grade Arithmetic
Ancillary Services	-.09	.13	.18	.33	.24	.19	.18	.16
Pupil Population in District	.12	.31	.12	.14	.15	.05	.00	-.02
Secondary Expenditures	-.11	-.05	-.08	-.07	-.07	-.08	-.07	-.21
% Non-Negro Pupils	.12	.09	.10	.00	.01	-.01	.20	.11
Community Prosperity	-.17	-.03	-.08	-.08	-.01	-.06	.15	.11
Elementary Expenditures	-.03	.00	.02	-.00	.08	.00	.14	.02
Noninstruct. Services	.12	.01	.02	.13	.05	.04	.07	-.04
% Non-Puerto Rican Pupils	-.07	-.03	.02	.03	.04	.02	.13	.03
Secondary Nongraded	-.01	-.03	.13	.02	.00	.07	.10	-.04
Innovative-ness	.00	.01	-.03	.06	.09	.10	.11	.08
School Enroll-ment	.10	-.01	.04	.03	.05	.10	-.05	-.08
Federal Resources	.01	.04	.06	.09	.00	.03	-.03	.00
Academic Facilities	.00	-.02	.05	.08	.00	-.03	-.07	.04
Nonacademic Facilities	-.07	.01	-.01	.00	-.01	.00	.02	.08
Uncrowdedness	-.02	-.03	.03	.07	-.06	.03	-.02	.00

Observations of interest in table 2 include the following:

1. Higher expenditures were, in general, associated with higher pupil achievement. However, in some expenditure categories the relationship tended to be reversed. Teachers' salaries were negatively related to all but 3 of the 8 indicators of pupil achievement. Guidance expenditures were negatively related to all but 1 of the 8 achievement categories. The number of professionals per pupil was negatively related to 4 of the 8 achievement categories.
2. With respect to the distribution of innovations and services within a district, it was found that these were positively correlated with nearly all of the achievement categories. In the case of guidance services, it was found that 2 of the 8 achievement categories were negatively correlated with the distribution of these services.
3. It appeared that the distribution of the following services or innovations were most highly associated with high achievement in a number of categories: educational TV, nongraded elementary programs, independent study programs, and miscellaneous local innovations.

TABLE 2.

CORRELATIONS BETWEEN ADMINISTRATIVELY CONTROLLABLE INPUT VARIABLES AND SELECTED OUTPUT VARIABLES (N=636)

	Output Variables							
Expenditures/Pupil	Propor. leaving to 2-year schools	Propor. passing Regents exam in bookkeeping	Propor. passing Regents exam in business law	Propor. passing Regents exam in shorthand	Propor. passing Regents exam in Latin	Propor. passing Regents exam in physics	Mean of third-grade PEP reading scores	Mean of ninth-grade PEP arithmetic scores
Pre-K	.00	-.06	-.06	-.02	.03	.03	.00	.00
Central Administration	.00	-.08	.02	.08	.54	.12	.03	.01
Salaries, K-6	.02	.13	.01	.06	.06	.05	.19	.00
Substitutes' Salaries	.06	.05	.08	.11	.24	.18	.34	.19
Equipment	.00	.00	.03	.03	.12	.07	.17	.05
Texts	.12	-.05	-.01	.01	.08	.11	.01	.02
Teaching	.03	-.04	-.05	-.03	-.07	-.01	.08	.03
Cocurricular Activities	.03	.00	.05	.04	.06	.38	.08	.04
Interscholastic Athletics	.04	.01	.04	.04	.04	.09	-.09	-.03
Guidance	-.02	-.07	-.01	-.04	-.04	.03	-.02	-.08
Total Instruc.(day sch.)	.03	.10	.03	.12	.08	-.02	-.08	.05
Instruction, Spec. School	-.05	-.05	-.02	-.04	-.01	-.02	.03	-.01
Professionals per Pupil	.01	-.05	-.03	-.02	.03	-.01	.45	.02
Proportion of Schools With								
Independent Study	.04	.12	.28	.24	.13	.17	.07	.04
E.T.V.	.12	.26	.22	.31	.19	.25	.07	.03
Nongraded Elem. Prog.	.10	.30	.31	.20	.33	.31	.18	.12
Local Innovations	.07	.04	.19	.15	.15	.18	.13	.04
Guidance Services	-.01	.06	.03	.04	.05	.05	.09	-.09
Health Services	.03	.12	.09	.13	.06	.05	.01	.00
Consultants	.00	.30	.12	.06	.02	.05	.01	.11

APPENDIX C

APPLICATION OF THE MODELS

Once the models are programmed for a computer, the user will be able simply to provide values for the sixth raw input variables, whereupon the computer will perform the calculations implied by the model equations

$$C = L \left\{ T_1 G \left\{ \frac{R_j - \mu_j}{\sigma_j} \right\} + B \right]_i \left[T_2 G \left\{ \frac{R_j - \mu_j}{\sigma_j} \right\} \right]_i \right\} + C_o$$

yielding estimated values for the eight output variables, together with appropriate information concerning the reliability of these estimates. The user will then provide information concerning changes in inputs, from which the models will estimate the behavior of the outputs--both the eight outputs that are explicitly in the equations and the twenty-two others that are associated with them in criterion clusters.

The actual matrices that constitute the parameters of the models are of course much too complicated to permit the demonstration of their operation without the aid of a computer. In order to show how the models work, therefore, we shall imagine a vastly simplified set of inputs, outputs, and parameters and trace through the operation of the model equations on these.

Let us suppose that we have only four raw input variables, whose values are 1, 2, 3, and 4 respectively; that their means are 0, 5, 2, 3, 2, and 5, respectively, and that their standard deviations are 0.5, 0.15, 0.5, and 1.0, respectively. The input matrices then have the following appearance:

---

*This section is an excerpt from the contracter's (Abt Associates) report.

$$R = \begin{bmatrix}1\\2\\3\\4\end{bmatrix} \qquad \mu = \begin{bmatrix}0.5\\2.3\\2.0\\5.0\end{bmatrix} \qquad \sigma = \begin{bmatrix}0.5\\0.15\\0.5\\1.0\end{bmatrix}$$

Suppose furthermore that the parameter matrices within the model are as follows:

$$G = \begin{bmatrix}1.00 & 0.00 & 0.50 & 0.00\\ 0.00 & -0.25 & 0.25 & 0.00\\ 0.00 & 1.00 & 0.00 & 2.00\end{bmatrix}$$

$$T_1 = \begin{bmatrix}0&0&0\\0&0&0\\0&0&0\\0&1&0\\0&0&1\end{bmatrix} \qquad T_2 = \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\1&0&0\\0&1&0\end{bmatrix} \qquad B = \begin{bmatrix}1\\1\\1\\0\\2.5\end{bmatrix}$$

$$L = \begin{bmatrix}0.500 & -0.400 & 0.330 & 1.500 & -2.000\\ -0.010 & -1.200 & -0.010 & 0.500 & 1.000\end{bmatrix} \qquad C_o = \begin{bmatrix}0.1\\-1.3\end{bmatrix}$$

Applying the model equations, we first compute the vector Z of standardized input data values:

$$Z = \left\{\frac{R_j - \mu_j}{\sigma_j}\right\} = \begin{Bmatrix}\frac{1-0.5}{0.5}\\ \frac{2-2.3}{0.15}\\ \frac{3-2}{9.5}\\ \frac{4-5}{1}\end{Bmatrix} = \begin{Bmatrix}1\\-2\\2\\-1\end{Bmatrix}$$

The function of the G matrix is to transform the four Z values into three index values Y; this operation is analogous to the dimensionality reduction accomplished by factor analysis and subsequent manipulation. The G matrix of our hypothetical example defines three indices as follows:

$$Y_1 = Z_1 + 0.5\ Z_3$$

$$Y_2 = -0.25\ Z_2 + 0.25\ Z_3$$

$$Y_3 = Z_2 + 2\ Z_4$$

Or, in matrix form:

$$Y = G\ Z = \begin{bmatrix} 1.00 & 0.00 & 0.50 & 0.00 \\ 0.00 & -0.25 & 0.25 & 0.00 \\ 0.00 & 1.00 & 0.00 & 2.00 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 2 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} (1 \times 1.00) + (-2 \times 0.00) - (2 \times 0.50) + (-1\ X\ 0.00) \\ (1 \times 0.00) + (-2\ X\ (-0.25)\ ) + (2 \times 0.25) + (-1 \times 0.00) \\ (1 \times 0.00) + (-2 \times 1.00) + (2 \times 0.00) + (-1 \times 2.00) \end{bmatrix}$$

$$= Y = \begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix}$$

Suppose now that we wish to define non-linear indices X such that

$$X_1 = Y_1$$

$$X_2 = Y_2$$

$$X_3 = Y_3$$

$$X_4 = Y_1\ \ Y_2$$

$$X_5 = Y_2\ (Y_3 + 2.5)$$

The matrices T_1, T_2, and B, as specified above, accomplish this transformation. Introducing, for convenience, the intermediate matrices A_1 and A_2 .

$$A_1 = T_1 Y + B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 2.5 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -4 \end{bmatrix} \quad + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 2.5 \end{bmatrix}$$

$$= A_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1.5 \end{bmatrix}$$

$$A_2 = T_2 Y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix}$$

$$= A_2 = \begin{bmatrix} 2 \\ 1 \\ -4 \\ 2 \\ 1 \end{bmatrix}$$

$$X = A_{1i} A_{2i} = \begin{bmatrix} 1 \text{ x } 2 \\ 1 \text{ x } 1 \\ 1 \text{ x } -4 \\ 1 \text{ x } 2 \\ -1.5 \text{ x } 1 \end{bmatrix}$$

$$= X = \begin{bmatrix} 2 \\ 1 \\ -4 \\ 2 \\ -1.5 \end{bmatrix}$$

Finally, the multiple non-linear regression coefficient matrix L defines the contribution of each non-linear predictor indices to the predictions of the two criterion variables C_1 and C_2. The intercept matrix simultaneously specifies what the predictions would be if all the predictors were zero. The final computations of this model operation take place as follows:

$$C = LX + C_o = \begin{bmatrix} 0.500 & -0.400 & 0.333 & 1.500 & -2.000 \\ -0.010 & -1.200 & -0.010 & 0.500 & 1.000 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -3 \\ 2 \\ -1.5 \end{bmatrix} + \begin{bmatrix} 0.1 \\ -1.3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -0.4 & -1 & +3 & +1 \\ -0.02 & -1.2 & +0.03 & +1 & -1.5 \end{bmatrix} + \begin{bmatrix} 0.1 \\ -1.3 \end{bmatrix}$$

$$= C = \begin{bmatrix} 3.70 \\ -2.99 \end{bmatrix}$$

The net result of these computations is to estimate values of 3.70 and -2.99 for the two criterion variables, given values of 1, 2, 3, and 4 for the four raw predictors. Suppose now that a school administrator proposes to make policy changes in his school district that would modify the predictor values from (1, 2, 3, 4) to (1.5, 2.15, 3, 3). The reader may wish to verify that the matrix operations demonstrated above yield criterion values of 8.159 and -1.640, when applied to this new set of input values. In this imaginary example, the values are meaningless, of course. But in a real school district, it might be very useful to compare the kinds of changes in educational outputs that can be expected to be associated with different possible configurations of inputs. The New York Education Department Models are designed to permit such comparisons.

6970/1/02303/9921